

# Index

## A

Absolute exponential (AE) kernel function, 172  
Activation function, 458  
Activation layer, 649, 650  
Adaptive Boosting (AdaBoost), 185  
Additive noise, 127  
Adjusted machine learning view  
  hypothesis space assumption, 387  
  Occam's Razor principle, 386  
  quality assurance, 386  
Aerial Image Measurement System (AIMS), 24–25  
Aging analysis, 10  
AlexNet, 657, 662  
A-miner, 619  
Analog and mixed-signal (AMS) circuits, 423  
  BMF (*see* Bayesian model fusion (BMF))  
  performance metrics, 403  
  post-silicon validation, 404  
  pre-silicon validation, 404  
Analysis of variance (ANOVA), 480  
Anchor gauges, 35  
Antecedent size, 620  
Antithetic variables, 327  
Approximate high-level taxonomy, 1–2  
Area under an ROC curve (AUC), 182, 197–198  
Artificial intelligence (AI), neural networks  
  backpropagation algorithm, 647, 648  
  CNN, 649–651  
  layer types, 649, 650  
  LSTMs, 651  
  resource requirements, 651–652

Artificial neural network (ANN), 98  
Assertion complexity, 621  
Assertion evaluation, 620–621  
Assertion expectedness, 620–621  
Assertion generation, 610  
  complexity analysis, 638  
  coverage guided association, 615  
  decision forest, 614–615  
  expectedness analysis, 637–638  
  importance analysis, 638–639  
  input space coverage and hit rate, 635–636  
  number of, 634  
  PRISM algorithm, 615  
  qualitative analysis, 638–641  
  quality, 614  
  RTL statement coverage, 637  
  runtime and memory usage, 636–637  
  scalability, 641  
  sequential depth, 635  
Assertion importance, 621  
Assertion mining algorithms, *see* GoldMine  
Association mining algorithm, 628  
Automated test equipment (ATE), 120  
Auto-regressive moving-average (ARMA) model, 255  
Autoscaling for multiple workloads  
  scaling down, 603–604  
  scaling up, 604–606  
Autoscaling for separate workloads, 602–603

## B

Bayesian inference, 131–132, 405  
Bayesian learning, 15

- Bayesian model fusion (BMF), 11–12
    - a priori distribution, 405
    - Bayesian inference, 405
    - post-silicon validation
      - Gaussian distribution, 418
      - likelihood function, 418
      - MAP estimation, optimization
        - formulation, 418–419
      - posterior distribution, 417
      - regression modeling, 416–417
      - VCO, 419–420
    - pre-silicon validation
      - distribution estimation, 411–416
      - moment estimation, 406–411
    - reuse, 405
  - Bayesian regularization, 486–487
  - Bayes' theorem, 209
  - Best-gain decision forest (BGDF) algorithm, 611–612, 625
    - analysis, 627
    - antecedent condition, 622–623
    - assertions, 626
    - decision forest, 623–624
    - example, 626–627
    - feature variables, 623, 626
    - input parameters, 624
    - optimizations, 623, 624
  - Bias temperature instability (BTI), 10, 265
  - Bias-variance trade-off, 465–467
  - Bit error rate (BER), 410
  - Black-box optimization
    - BO based on GP, 509–511
    - machine learning, 508–509
    - TSBO
      - acquisition function, 514–515
      - algorithm progression times, 517
      - Euclidean distance, 513
      - fast exploration stage, 512
      - hierarchical partitioning scheme, 514
      - IMGPO and GP-UCB, 516
      - learning acquisition functions, 513
      - log distance to optima, 516
      - in MATLAB, 515, 516
      - pure exploration stage, 514
      - run time statistics, 518
  - Blockade filtering, 308
  - BMF, *see* Bayesian model fusion
  - Boolean learning, 388
  - Boolean satisfiability, 389
  - Built-in proactive tuning (BIPT) system, 269
  - Built-in self-test (BIST), 425
    - charge-pump PLL, 440
    - classification techniques, 441
    - classifier performance comparison, 441
    - CP and VCO quiescent currents, 441
    - scheme synthesis, 440
    - signature ranking, 442
    - test time optimization, 442
- C**
- CDF, *see* Cumulative distribution function
  - Cell failure probability, extreme statistics
    - chip/array failure probability, 293
    - failure threshold, 293
    - incorporating redundancy, 295–296
    - in parts-per-billion (ppb), 295
    - random dopant fluctuation, 294
    - units, 294–295
  - Center for Advanced Electronics through Machine Learning (CAEML), 15
  - Central limit theorem (CLT), 357, 358
  - Chip disposition criteria, 245
  - Chip-level silicon characterization, 220
  - Chip slack, 246
  - Chip testing, 9
  - CNV accelerator, 663
  - Compact lithographic process models, 7
  - Compact network architectures, 656–657
  - Compact process model (CPM)
    - machine learning
      - definition, 29
      - model tasks, 29–31
      - performance metrics, 33–35
      - physical mechanisms, 27–28
      - rigorous/TCAD simulators, 28
      - supervised learning (*see* Supervised learning)
      - training experience, 31–33
    - neural network (*see* Neural networks (NNs))
  - Compact set, 460
  - Compressed sensing, 203
  - Concentric circle area sampling (CCAS), 97–98
  - Confidence interval (CI)
    - scaled-sigma sampling, 370, 371
    - subset simulation, 357, 361, 362
  - Constrained process discovery approach, 396
  - Control variates method, 327
  - Convolutional neural network (CNN), 649–651
    - AlexNet architecture, 58–60
  - Convolution (CONV) layer, 649
  - Core scaling and thread assignment
    - autoscaling for multiple workloads, 603–606
    - autoscaling for separate workloads, 602–603

- DRAM power, 599–600
  - energy-aware workload execution, 600–602
  - flow diagram, 599
  - PARSEC workloads, 598
- Covariance function, 125
- Covariance matrices, 169
- Coverage guided mining algorithm, 612, 628–631
- CPM, *see* Compact process model
- Critical dimensions (CDs), 74–75, 82, 91
- Cross-validation, 381, 487–488
- Crude Monte Carlo (CMC) technique, 350
- C-Struts algorithm
  - auto-regressive moving-average model, 255
  - auxiliary probability distribution, 255, 258–259
  - column cluster prototype, 257
  - conditionally independent, 256
  - conditional probability, 255
  - joint probability, 257–258
  - marginal probability, 255
  - multiway clustering algorithms, 258
  - optimal mapping functions, 258
  - posterior distribution, 256
  - row cluster prototype, 257
  - time series, 254
- Cumulative distribution function (CDF), 296–297
  
- D**
- Data collection platform, LAR
  - modeling methodology, 580
  - mode specific registers, 579
  - PARSEC, 580
  - RAPL interface, 580
  - RHLE operating system, 579
  - SPEC CPU benchmark, 580
  - voltage regulation module, 579
- Dataflow centric optimizations, 671
- Data mining algorithms, 612, 616
- Datapath-aware placement (PADE), 107–108
- Data retention voltage (DRV), 313–315, 319–321
- DCT, *see* Discrete cosine transform
- Decision engine algorithms, 547–548
  - base decision algorithm, 549–550
  - comparisons, 556–559
  - 14 nm data mining, 555
  - IBM server chips, 554
  - learning algorithm, 551–553
  - z13 22 nm server processor, 554–555
- Decision tree-based learning, 621–622
- Decision tree induction algorithm, 111
- Deep learning, 378
- Deep neural networks (DNNs), 51, 98
  - energy-efficient design, 647, 652, 658–659, 662
  - high-dimensional layout, 99
  - manual trial and error process, 100
  - MGD algorithm, 100
  - operation of convolution, 99
  - ReLU layer, 99–100
  - state-of-the-art hotspot detectors, 101
- Delay monitoring schemes, 265
- Dense line and space (DLS), 80
- Dense set, 460
- Deposition, 21
- Design for test (DfT) infrastructure, 267
- Design space exploration (DSE) approach, 539
- Design-time solutions, 265
- Detection accuracy, 96
- Die-to-die (D2D) variation, 237
- Diffraction-limited system, 24
- Dirac delta function, 208
- Discrete cosine transform (DCT), 9, 122–123
  - flush delay measurement data, 221, 222
  - leakage current measurement data, 226, 227
  - MAP estimation, 206–207
  - preselection, 216–217
  - RO period measurement data, 228, 229
- Discriminative model, 4
- Dissolution inhibitor concentration, 53
- DNNs, *see* Deep neural networks
- DRV, *see* Data retention voltage
- Dual-lane data communication, 443
- Dynamic reliability management (DRM), 269
- Dynamic voltage and frequency scaling (DVFS), 505, 572
  
- E**
- Edge placement error (EPE) tolerance, 70
- EIE, 663
- Electronic design automation (EDA), 245–247
- Empirical error rate, 383–384
- Energy-efficient design
  - challenges
    - accuracy vs. energy trade-off, 673
    - adaptability, (re-)configurability, and scalability, 673
    - benchmarking with fairness and high fidelity, 674
    - neuromorphic computing, 672–673
    - open-source contributions, 674
    - optimizing memory vs. computations, 672

- Energy-efficient design (*cont.*)
    - run-time evolutionary algorithms, 674
  - error resilience analysis
    - data distributions, 667
    - DNN architecture, 664
    - feature maps, 665, 666
    - IMPACT, 666
    - LeNet-5, 664
    - NNs training, 664
  - hardware-level architectures
    - accelerators, dataflows and architectures, 658–659
    - deep CNN, 660–661
    - FPGA/ASIC-based systems, 658
    - leveraging sparsity, 662–663
    - memory-efficient architectures, 661–662
  - neural networks
    - backpropagation algorithm, 647, 648
    - CNN, 649–651
    - effective fixed-point quantization, 670
    - H/W and architectural optimizations, 670–671
    - layer types, 649, 650
    - LSTMs, 651
    - model compression, 668–670
    - resource requirements, 651–652
    - training, 668
  - software and co-design optimizations
    - compact network architectures, 656–657
    - vs. hardware, 657–658
    - pruning, 653–655
    - ResNet50, 652
    - weight sharing, 655–656
  - Equivalent sigma ( $\sigma$ ), 295
  - Error resilience analysis
    - data distributions, 667
    - DNN architecture, 664
    - feature maps, 665, 666
    - IMPACT, 666
    - LeNet-5, 664
    - NNs Training, 664
  - Etch bias model, *see* Machine learning-guided EPC (ML-EPC)
  - Etching, 21
  - Etch proximity correction (EPC), 7, 61
  - Expectation maximization (EM), 122, 436
  - Expectedness analysis, 637–638
  - Extreme statistics
    - cell failure probability
      - chip/array failure probability, 293
      - failure threshold, 293
    - incorporating redundancy, 295–296
    - in parts-per-billion (ppb), 295
    - random dopant fluctuation, 294
    - units, 294–295
  - statistical blockade, sampling rare events algorithm, 309
  - blockade filtering, 308
  - circuit performance metric, 307
  - conditionals and disjoint tail regions, 313–316
  - efficient tail sampling method, 309, 310
  - extremely rare events and their statistics, 316–317
  - recursive formulation, 317–321
  - sampling and commercially available device models, 310–311
  - 6T SRAM Cell, 311–313
  - tail and body regions, 308
  - unbiasing classifier, 309
  - tails (*see* Tails)
- Extreme statistics in memories, 10
- Extreme value theory (EVT), 298
- F**
- False alarm, 96
  - False positive rate (FPR), 181, 198
  - Fast convolution unit (FCU), 660
  - Fast statistical analysis
    - Bernoulli trial, 325
    - importance sampling method, 327–329
    - logistic regression, 10 (*see also* Logistic regression-based fast statistical analysis)
    - memory design, 323–324
  - Monte Carlo method
    - Buffon's needle experiment, 325
    - fail probability, circuit design, 326–327
  - rare circuit failures, 11
  - state-of-the-art FinFET SRAM design
    - experimental setup, 341
    - importance sampling stage, 343, 344
    - selective boosting and write-assist circuitry, 338–341
    - uniform sampling stage, 342, 343
    - yield estimation and convergence analysis, 343–346
    - statistical inference, 325
    - variance reduction method, 327
  - Feature-based analytics, 378
  - Feature-generating function, 40
  - Feature selection, 378

- Flush delay measurement data
  - DCT coefficients, 221, 222
  - measured flush delay values, 221, 222
  - M-LHS, 222
  - sparse pattern, 221
  - spatial sample generation, 222–224
  - spatial variation prediction, 224–226
- 14 nm data mining, 555
- Fully connected (FC) layer, 649
- Functional tests, 375, 376
  
- G**
- Gaussian distribution
  - post-silicon validation, 418
  - pre-silicon validation
    - distribution estimation, 411–413
    - moment estimation, 406–408
- Gaussian process models
  - e-test measurements
    - leave-one-out cross-validation, 147
    - measurement site diagram, 145–146
    - parameters, 145
    - prediction errors, 148–150
    - sampling rates, 146
    - virtual probe, 147–148, 151
  - HVM yield estimation
    - absolute percentile error metric, 162
    - baseline spatial correlation model, 161
    - correlation model enhancement, 162–164
    - mean error, 164–165
    - parametric probe-test measurements, 161
    - prediction error, 164
    - prediction models, 161–162
    - results, 165–168
  - IC parameters
    - actual and predicted lengths, 158–160
    - current value vs. trimmed time, 157–158
    - laser trimming machine, 156–157
    - length-based original-target prediction method, 158–161
    - rate-based original-target prediction method, 158–161
  - probe-test specification measurements
    - discontinuous spatial patterns, 152–153
    - prediction errors, 152–155
    - Test Escape, 154, 156
    - Yield Loss, 154, 156
  - regression models (*see* Linear regression models)
    - for wafer-level correlations, 8
    - wafer-level test measurements
      - ATE, 120
      - e-test measurements, 120
      - specification tests, 120
      - statistical correlations, 120–123
- Gauss–Newton method, 485
- Generalized linear mixed model (GLMM), 106
- Generalized Pareto distribution (GPD), 299
  - mean excess function, 304
  - parameters
    - maximum likelihood estimation, 305–306
    - probability-weighted moment matching, 306–307
    - statistical blockade sampling method, 305
    - probability density function, 300
- Generative adversarial networks (GANs), 4–5
- “Golden” failure rate
  - scaled-sigma sampling, 371
  - subset simulation, 360–361
- GoldMine
  - A-miner, 619
  - assertion evaluation, 620–621
  - assertion generation, 610, 614–616
    - complexity analysis, 638
    - expectedness analysis, 637–638
    - importance analysis, 638–639
    - input space coverage and hit rate, 635–636
    - number of, 634
    - qualitative analysis, 638–641
    - RTL statement coverage, 637
    - runtime and memory usage, 636–637
    - scalability, 641
    - sequential depth, 635
  - BGDF algorithm, 611–612, 625
    - analysis, 627
    - antecedent condition, 622–623
    - assertions, 626
    - decision forest, 623–624
    - example, 626–627
    - feature variables, 623, 626
    - input parameters, 624
    - optimizations, 623, 624
  - coverage guided mining algorithm, 612, 628–631
  - data generator, 619
  - data mining algorithms, 612
  - decision tree algorithms, 611
  - decision tree-based learning, 621–622
  - dynamic analysis, 611–612
  - formal verification, 619
  - input and output event, 613
  - IODINE analyzes, 613

GoldMine (*cont.*)  
 methodology, 610  
 principle, 616–617  
 PRISM algorithm, 612, 631–633  
 RTL, 610  
 static analysis, 611–612, 618–619  
 system-on-a-chip, 609  
 Verilog designs, 633  
 GoldMine system, 14  
 GoogLeNet, 657  
 Greedy set covering algorithm, 628

**H**

HBM-based approach, 106–107  
 HEAR algorithm, 249–252  
 Hessian matrix, 332  
 Hierarchical Bayesian model (HBM), 71, 106  
 High-volume manufacturing (HVM)  
 density estimation approaches, 141  
 histogram-based estimation method,  
 141–142  
 KDE, 142–144  
 post-silicon design validation, 141  
 prediction evaluation, 144–145  
 yield estimation  
 absolute percentile error metric, 162  
 baseline spatial correlation model, 161  
 correlation model enhancement,  
 162–164  
 mean error, 164–165  
 parametric probe-test measurements,  
 161  
 prediction error, 164  
 prediction models, 161–162  
 results, 165–168  
 Histogram with Random Sampling (Hist-RS),  
 166–167  
 Histogram with Spatiotemporal GP and  
 Progressive Sampling (Hist-GP-ST-  
 PS), 166–167  
 History-based enhancement  
 jump-start algorithm, 559–561  
 recommender system, 561–562  
 Hit rate, 620, 635–636  
 Hyperbolic tangent function, 474  
 Hypothesis space, 388–390

**I**

Importance sampling (IS) methods, 350  
 Gaussian distribution, 328  
 integrated importance sampling, 328–329  
 natural distribution, 327, 328

ratioed importance sampling, 328  
 regression importance sampling, 328  
 Input-referred correlation (IRC), 480–481  
 Input space coverage, 620, 635–636  
 Integrated circuits (ICs)  
 actual and predicted lengths, 158–160  
 application, 176–177  
 concept drift, 176  
 current value vs. trimmed time, 157–158  
 flash memory chip manufacturing, 177–178  
 imbalanced classification, 175–176  
 KGD procedures, 177  
 laser trimming machine, 156–157  
 learning models (*see* Learning models)  
 length-based original-target prediction  
 method, 158–161  
 packaging and memory testing process  
 with classifier, 179  
 confusion matrices, 180–181  
 mathematical formulation, 182–184  
 ROC curve, 181–182  
 without classifier, 179  
 rate-based original-target prediction  
 method, 158–161  
 Integrated importance sampling, 328–329  
 Integrated voltage regulator optimization,  
 12–13  
 Isolated line (Iso), 80  
 Iterative feature search  
 dataset construction, 379  
 domain knowledge, 380, 396–397  
 feature selection, 379  
 feature set, 380  
 model evaluation step, 381  
 process, 379  
 sample selection, 379  
 tool requirement, 381–382

**J**

Joint space–time sampling, 283–284  
 Jump-start algorithm, 559–561

**K**

KDE with Random Sampling (KDE-RS),  
 166–167  
 KDE with Spatiotemporal GP and Progressive  
 Sampling (KDE-GP-ST-PS),  
 166–167  
 Kernel-based methods, 456  
 Kernel density estimation (KDE), 142–144  
 Kernel function, 125  
 Kernel weighting

- atomized kernel, 426–430
  - EDA domain, 425
  - learning model, 430–431
  - mapped space, 429
  - Mercer’s condition, 430
  - methods, 425–426
  - noise/redundant features, 427
  - radial basis function, 427, 430
  - 6-dimensional space, 427
  - support vector machines, 425
- Known good die (KGD) test, 177
- Kolmogorov–Smirnov (KS) test, 167–168
- Kronecker delta function, 127
  
- L**
- Labeled data, 31
- Latent variable regression (LVR), 12, 451, 455–457
- Latin hypercube sampling (LHS), 214–216
- Leakage current measurement data, 226–228
- Learning from limited data
  - coverage point, 375, 376
  - dataset format, 375, 376
  - functional verification, application in, 375, 376
  - iterative feature search (*see* Iterative feature search)
  - layout snippets, 376, 377
  - learning about features, 378–379
  - limited number of samples, 378
  - negative samples, 375
  - positive samples, 375
  - SAT-based implementation
    - Boolean learning, 388
    - hypothesis space pruning, 389–390
    - monomial learning, 388–389
    - SAT encoding, 390–391
    - uniqueness requirement, 391–396
  - STA-critical path, 377
  - 350 critical paths not predicted by STA, 376–377
- Learning models
  - class imbalance, 189
  - concept drift
    - AUC results, 197
    - example of, 188–189
    - issue of, 188
    - optimal threshold, 196
    - probability distribution, 187–188
    - ROC curves, 195–196
    - semiconductor manufacturing, 188
    - substantial changes, 189
  - decision tree, 191
  - imbalanced classification problem, 185
  - incremental learning, 197–198
  - Learn+ +.NIE algorithm, 190–191
  - from limited data, 11
  - online machine learning
    - effectiveness, 193–195
    - good classifier, 179, 186
    - Online RUSBoost, 187–188
    - test instances, 186
    - traditional batch learning, 187
  - RUSBoost
    - AdaBoost framework, 185–186
    - imbalanced dataset, 192–193
    - random undersampling, 186
- Least absolute shrinkage and selection operator (LASSO), 574
- Least-angle regression
  - advantages, 572
  - algorithm, 576, 577
  - core scaling and thread assignment
    - autoscaling for multiple workloads, 603–606
    - autoscaling for separate workloads, 602–603
    - DRAM power, 599–600
    - energy-aware workload execution, 600–602
    - flow diagram, 599
    - PARSEC workloads, 598
  - correlation coefficients, 578
  - data collection platform
    - modeling methodology, 580
    - mode specific registers, 579
    - PARSEC, 580
    - RAPL interface, 580
    - RHLE operating system, 579
    - SPEC CPU benchmark, 580
    - voltage regulation module, 579
  - Efron’s seminal work, 575
  - lasso, 574
  - mathematical and algorithmic background, 573
  - power proxies
    - in CMOS, 581
    - multicore case, 583–586
    - single-core case, 582–583
  - regression coefficients, 579
  - ridge regression, 574
  - sleep slates, 591–594
  - statistical metric model, 572
  - temperature proxies
    - multicore case, 589–591
    - single-core case, 586–588
  - workload signature, 594–597

- Leave-one-out cross-validation error, 529
  - Levenberg–Marquardt (LM) algorithm, 483–486
  - Linear (LI) kernel function, 171
  - Linear models, RSM, 450, 452–453
  - Linear regression models
    - covariance function value, 130
    - covariance with kernel functions, 124–126
    - discontinuous effect
      - Calinski and Harabasz index, 136–137
      - $k$ -Means clustering algorithm, 136–137
      - overview of, 134–135
      - partitioning procedure, 135–136
      - spatial patterns, 133
      - wafer measurement data, 134
    - Gaussian random fields, 123
    - GP-ST-PS, 140
    - HVM
      - density estimation approaches, 141
      - histogram-based estimation method, 141–142
      - KDE, 142–144
      - post-silicon design validation, 141
      - prediction evaluation, 144–145
    - hyperparameters
      - Bayesian inference, 131–132
      - kernel functions, 130–131
      - Markov chain Monte Carlo, 132–133
      - MLE, 133
      - model selection and adaptation procedure, 132
      - multivariate Gaussian distribution, 132
      - SA algorithm, 133
    - kernel stationarity property, 130–131
    - length scale parameter, 128–129
    - modeling radial variation, 127–129
    - monolithic linear regression formulation, 123
    - overview of, 123–124
    - predictive distribution, 126
    - production test deployment, 140–141
    - progressive sampling, 137–139
    - regularization parameter
      - additive noise, 127
      - decision-theoretic empirical risk minimization, 126
      - effects, 127–129
      - hyperparameters, 129–130
    - spatiotemporal feature inclusion, 138–140
    - training locations, 126
  - Line-end to line-end (E2E) patterns, 80
  - Lithography
    - CPM (*see* Compact process model (CPM))
    - patterning process
      - deposition/etching, 21
      - distributions, 23
      - etch transfer function, 26–27
      - image transfer function, 22, 24–25
      - machine learning technology, 21
      - mask transfer function, 22–24
      - Moore’s law, 20–21
      - resist transfer function, 22, 25–26
      - schematic of, 27–28
      - shrinking device size, 21
      - silicon wafer, 21
  - Lithography compliance check (LCC), 101–102
  - Local clock buffer (LCB), 110
  - Logistic regression
    - hypothesis function, 331
    - likelihood function, 331
    - numerical optimization techniques, 332–333
    - simple review, 329–330
  - Logistic regression-based fast statistical analysis
    - cross-validation, 337
    - methodology flow diagram, 334, 335
    - regularization framework, 335–337
    - terminology and definitions, 333–334
  - Logistic regression-based importance sampling methodology, 323–324
  - Log-likelihood function, 305
  - Long short-term memory (LSTMs) networks, 651
  - Lookup tables (LUTs), dynamic adaptation, 269
  - Low-dropout regulator (LDO), 425
    - regression performance, 438
    - variability analysis, 437–440
  - Lumped model, 27
  - LVR, *see* Latent variable regression
- M**
- Machine learning (ML)
    - adjusted machine learning view
      - hypothesis space assumption, 387
      - Occam’s Razor principle, 386
      - quality assurance, 386
    - assumptions in, 382–383
    - and big data analytics, 233–234
    - clustering, 235
    - co-clustering structural temporal data
      - C-Struts algorithm (*see* C-Struts algorithm)
    - experimental results, 259–260



- problem formulation, 253–254
  - time series data, 253
- and data mining, 234
- definition, 29
- fast statistical analysis (*see* Fast statistical analysis)
- mask synthesis (*see also* Machine learning-guided OPC (ML-OPC))
  - OPC, 104–107
  - SRAFs, 102–104
  - standard flow, 101–102
- model tasks, 29–31
- MTL problem, 235
- performance metrics, 33–35
- physical design, 106
  - clock optimization, 110–111
  - datapath placement, 107–108
  - lithography routing, 111–112
  - routability-driven placement, 108–110
- physical mechanisms, 27–28
- physical verification
  - example of, 96–97
  - hotspot detection, 98–101
  - layout feature extraction, 97–98
  - lithography, 96
  - model selection, 97–98
- procedure of, 69–70
- regression problem, 234–235
- rigorous/TCAD simulators, 28
- robust spatial correlation extraction
  - algorithms, 239–243
  - experimental results, 243–244
  - problem formulations, 237–239
  - technology scaling, 236
- statistical chip testing and yield optimization
  - chip disposition criteria, 245
  - statistical test margin computation, 245–247
  - surrogate metrics, 245
  - target metrics, 245
- supervised learning (*see* Supervised learning)
- time series data, 235
- traditional
  - empirical error rate, 383–384
  - Occam's Razor, 385
  - overfitting, 383–386
  - underfitting, 384
- training experience, 31–33
- unsupervised learning technique, 235
- virtual metrology, 235
- wafer quality prediction
  - experimental results, 252–253
  - HEAR algorithm, 249–252
  - problem formulation, 247–249
- Machine learning-based aging analysis
  - aging prediction method
    - benefits of, 267
    - proposed technique (*see* Offline correlation analysis; Runtime stress monitoring)
  - $DD_{\min}$ , 270
  - delay monitoring scheme, 265–266
  - DRM techniques, 269
  - experiments
    - area and power overhead, 286
    - Leon3 and OpenRISC 1200 (OR1200), 281
    - overhead at design time, 286
    - performance overhead, 285–286
    - prediction accuracy evaluation, 283–284
    - setup, 282
    - SVM training and validation, 282
    - time-sampling hardware design validation, 285
  - flip-flops, 271
  - LUTs, dynamic adaptation, 269
  - negative bias temperature instability, 267–268
  - online circuit failure prediction, 269
  - proactive aging mitigation, 266
  - signal probabilities, 266–267
  - state-of-the-art monitoring methods, 270
  - time constant, 266
  - workload-induced stress monitoring, 266
  - worst-case guard bands, 268–269
- Machine learning (ML)-based Bayesian optimization (BO), 505
- Machine learning-guided EPC (ML-EPC), 78
  - construction of, 84–86
  - EPC algorithm, 86–87
  - etching process, 79–80
  - factors, 81–82
  - litho pattern, 81–82
  - MB-EPC, 81
  - parameter extraction, 83–84
  - RB-EPC, 80
  - training segments
    - assessment of accuracy, 88
    - CD errors, 91
    - effect of cross-validation, 90–91
    - parameter sets, 90
    - preparation, 80, 82–83
    - RMS errors, 89
    - sampling of, 90

- Machine learning-guided OPC (ML-OPC)  
 accuracy and runtime, 71–72  
 EPE tolerance, 70  
 parameters, 72–73  
 PFT signal  
 assessment of parameters, 76–77  
 basis function, 73–74  
 benefits, 75  
 Bessel function, 73  
 complexity of illumination, 77  
 critical dimensions, 74–75  
 hybrid ML-and MB-OPC, 77–78  
 intensity of light, 74  
 layer and device, 77–78  
 layout image, 74  
 MLP construction, 75–76  
 testing phase, 70–71  
 training phase, 70–71
- Markov chain Monte Carlo (MCMC), 11,  
 132–133, 355–358
- Mask diffraction model, 50–52
- Mask rule check (MRC), 101–102
- Mask synthesis problem, 7–8
- Master–slave flip-flop (MSFF), 490–492
- Maximum a posteriori (MAP) estimation, 203,  
 204
- Maximum-likelihood estimation (MLE), 133,  
 207  
 scaled-sigma sampling, 366, 369  
 tail estimation, GPD parameters, 305–306
- MCMC, *see* Markov chain Monte Carlo
- Mean excess function, 303–304
- Mean squared error on the training set  
 ( $MSE_{\text{train}}$ ), 43
- Memory centric optimizations, 671
- Memory-efficient architectures, 661–662
- Mini-batch gradient descent (MGD) algorithm,  
 100
- MNIST dataset, 664
- MobileNets, 656
- Model-based (MB) approach, 106–107
- Model-based EPC (MB-EPC), 81
- Model-based OPC (MB-OPC), 26  
*see also* Machine learning-guided OPC
- Model response  
 sensitivity analysis, 533–534  
 storage model, 530–533
- Mode specific registers (MSRs), 579
- Modified-Latin hypercube sampling (M-LHS)  
 algorithm, 214–216, 222
- Modified Metropolis (MM) algorithm, 351
- Monomial learning, 388–389
- Monte Carlo (MC) method, 243, 506  
 Buffon’s needle experiment, 325  
 fail probability, circuit design, 326–327
- Multicore processors, 14
- Multilayer perceptron (MLP), 71, 75–76
- Multi-macro tuning, STSS  
 priority ranking, 565–566  
 second pass SynTunSys runs, 566  
 STSS results, 566–567
- Multi-objective optimization of IVR  
 buck converter efficiency model, 520–521  
 convergence comparison, 526  
 embedded inductor, 521–522  
 GP-UCB and IMGPO, 525  
 hand-tuned design, 524  
 non-ML and ML techniques, 519  
 optimization setup, 522–523  
 power efficiencies, 524  
 Type I inductors, 524
- Multitask learning (MTL) problem, 235
- Multitask learning (MTL) problem, wafer  
 quality prediction  
 experimental results, 252–253  
 HEAR algorithm, 249–252  
 problem formulation, 247–249
- Multivariate Gaussian distribution, 132,  
 170–171
- ## N
- Negative bias temperature instability (NBTI),  
 267–268
- Neural networks (NNs)  
 backpropagation algorithm, 647, 648  
 CNN, 649–651  
 computational architecture, 47–48  
 design methodology  
 effective fixed-point quantization, 670  
 H/W and architectural optimizations,  
 670–671  
 model compression, 668–670  
 training, 668  
 etch transfer functions, 61–63  
 image transfer function, 50–52  
 layer types, 649, 650  
 LSTMs, 651  
 machine learning, 63–66  
 mask transfer function, 48–49  
 resist model  
 CNN, 58–60  
 development rate, 53  
 dissolution inhibitor concentration, 53  
 human intervention and computation,  
 53  
 Mardiris’ model, 53–54  
 performance metric, 52

- three-dimensional resist model, 55–58
  - threshold, 52
  - training data, 53–54
  - two-dimensional resist model, 55–58
  - Zach's model, 41, 54–55
  - resource requirements, 651–652
  - training algorithms, 47–48
  - Newton's method, 483, 484
  - NN/application-specific optimizations, 670–671
  - No-free-lunch (NFL) theorem, 381
  - Nonparametric methods, 4
  - Nonstationary kernel functions, 130
  - Normal distribution, 170–171
  - Normalization layer, 649, 650
  - Normalized mean square error (NMSE), 438
  - Novel hardware architectures, 14
  - Nyquist–Shannon sampling theorem, 202, 221
- O**
- Occam's Razor, 385
  - Occam's Razor principle, 386
  - Offline correlation analysis
    - and prediction model generation
      - aging-induced delay degradation and SP extraction, 272–273
      - design time, 271
      - offline-characterization phase, 272
      - predictor training using support-vector machines, 273–275
    - representative flip-flop selection, 271
      - correlation-based flip-flop selection, 275–276
      - fan-in cone-based flip-flop selection, 277–278
      - time complexity, 278–279
    - time sampling, 279–280
  - Online machine learning
    - effectiveness, 193–195
    - good classifier, 179, 186
    - Online RUSBoost, 187–188
    - test instances, 186
    - traditional batch learning, 187
  - Optical proximity correction (OPC), 29, 79–80
    - effectiveness, 104–105
    - flow of, 105–106
    - HBM, 106–107
    - MB approach, 106–107
    - problems, 7
  - Optimization oriented design flow, 506–507
  - Overall chip variance, 238
  - Overall detection and simulation time (ODST), 101
  - Overfitting, 383–386
  - Overlap and Add method (OaA), 660
- P**
- Parametric methods, 4
  - Parts-per-billion (ppb), 294, 295
  - Parts-per-million (ppm), 294
  - Pass gate device (PGL), 341
  - Performance counter monitor (PCM), 580
  - Performance of interest (PoI), 351
  - Phase-locked loop (PLL), 425
  - Photolithography
    - etch transfer function, 26–27
    - image transfer function, 22, 24–25
    - mask transfer function, 22–24
    - resist transfer function, 22, 25–26
  - Physical design, 8
  - Physical verification, 7–8
  - Polar Fourier transform (PFT) signals
    - assessment of parameters, 76–77
    - basis function, 73–74
    - benefits, 75
    - Bessel function, 73
    - complexity of illumination, 77
    - critical dimensions, 74–75
    - hybrid ML-and MB-OPC, 77–78
    - intensity of light, 74
    - layer and device, 77–78
    - layout image, 74
    - MLP construction, 75–76
  - Polynomial chaos expansion
    - experimental design, 527
    - LARS algorithm, 528
    - leave-one-out cross-validation error, 529
    - post-processing, 530
    - from standard to sparse, 528–529
  - Polynomial (POLY) kernel function, 171
  - Pooling layer, 649, 650
  - Post-exposure bake (PEB), 39
  - Post-silicon tuning, 221
  - Post-silicon validation
    - BMF
      - Gaussian distribution, 418
      - likelihood function, 418
      - MAP estimation, optimization
        - formulation, 418–419
        - posterior distribution, 417
        - regression modeling, 416–417
        - VCO, 419–420
      - challenges for, 404
  - Power delivery network (PDN), 505
  - PPR, *see* Projection pursuit regression
  - Prediction error, 217

- Pre-silicon validation
    - BMF
      - distribution estimation, 411–416
      - moment estimation, 406–411
    - challenges for, 404
  - PRISM algorithm, 612, 615, 631–633
  - Probability density function (PDF), 207, 208, 506
  - Probability-weighted moment (PWM)
    - matching, 306–307
  - Process design kit (PDK), 443
  - Processing elements (PEs), 658–659
  - Product quality loss (PQL), 245
  - Progressive sampling (PS), 137–139
  - PROjection-Based Extraction (PROBE), 454–455
  - Projection pursuit regression (PPR), 12
    - bias-variance trade-off, 465–467
    - convergence
      - best function  $g_i(t)$ , 469–472
      - “greedy” approach, 467
      - $i$ -th iteration, 468
      - kernel function, 472, 473
      - mean squared error, 469
      - nonidealities, 468
      - probability distribution, 467–468
      - projection vector  $w$ , 469–472
    - definition, 457
    - LVR, 458
    - neural networks, 457–458
    - smoothing method, 465–467
    - 3LP, 458
  - Proxies, 14
  - Pruning, 653–655
- Q**
- Quadratic model, 453–454
  - Quantiles
    - binomial distribution, 303
    - empirical distribution, 302–303
    - mean excess plot, 303–304
    - normal distribution, 303
    - sample quantile, 302
- R**
- Radial basis function (RBF), 427, 430
  - Radius basis function (RBF), 108
  - Random dopant fluctuation (RDF), 490
  - Rare circuit failure events
    - CMC, 350
    - failure boundary, 350
    - IS methods, 350
    - SRAM bit-cell, 349, 350
    - SSS approach (*see* Scaled-sigma sampling (SSS))
    - SUS technique (*see* Subset simulation (SUS))
  - Ratified importance sampling, 328
  - Raw probability (P), 294
  - Reaction–diffusion (RD) model, 267, 268
  - Receiver operating characteristic (ROC) curve, 181–182, 195–196
  - Rectified linear unit (ReLU), 662
  - Red Hat Linux Enterprise (RHLE), 579
  - Reduced-rank quadratic model, 454–455
  - Reduced-rank regression (RRR), 456
  - Register transfer level (RTL) design, 610, 637
  - Register transfer level (RTL) designs, 14
  - Regression importance sampling, 328
  - Relevance vector machine (RVM), 424
  - Response surface models (RSM)
    - circuit performance-metric/output space, 451–452
    - dimensionality, 450
    - linear models, 450, 452–453
    - LVR, 451, 455–457
    - nonlinearity, 450
    - PPR (*see* Projection pursuit regression (PPR))
    - PROBE, 454–455
    - quadratic model, 453–454
    - ridge functions
      - definition, 457, 458
      - degree of approximation, 463–464
      - density, 461–462
      - terminology, 460
      - univariate ridge functions, 459
    - SiLVR (*see* SiLVR model)
    - SPICE simulations, 452
    - variations, 450
  - Ridge functions
    - definition, 457, 458
    - degree of approximation, 463–464
    - density, 461–462
    - terminology, 460
    - univariate ridge functions, 459
  - Ring oscillator (RO) period measurement data, 228–230
  - Robust spatial correlation extraction
    - algorithms
      - global variation component, 243
      - homogeneous and isotropic random field, 240
      - inter-chip global variation, 241
      - least-square minimization problem, 242
      - overall process correlation, 240

- overall within-chip variance, 240
  - process variation, characterization, 242, 243
  - regression-like optimization problem, 241, 242
  - spatial variation component, 242
  - unbiased sample covariance, 241
  - unbiased sample variance, 240–241
- experimental results, 243–244
- problem formulations, 237–239
- technology scaling, 236
- ROC curve, *see* Receiver operating characteristic curve
- Root-mean-square error (RMSE), 34, 77
- Roughness penalty, 486
- RSM, *see* Response surface models
- Rule-based EPC (RB-EPC), 80
- Running Average Power Limit (RAPL) interface, 580
- Runtime stress monitoring, 271, 280–281
- S**
- Sample mean excess plot, 303
- SAT encoding, 390–391
- Scaled-sigma sampling (SSS), 351
  - confidence interval, 370, 371
  - covariance matrix, 367
  - first-order optimality condition, 369
  - Gaussian distribution, 367, 370
  - “golden” failure rate, 371
  - high-probability failure region, 363
  - hyper-rectangles, 364–366
  - independent Monte Carlo simulations, 367
  - linear regression, 370
  - mean vector, 368
  - MLE solution, 366, 369
  - model coefficients, 370
  - original failure rate, 364, 366
  - pre-selected scaling factors, 369
  - probability distribution, 367
  - scaled failure rate, 363–364, 366–368
  - scaled PDF, 362
  - scaling factor, 362, 366
  - soft maximum, 365
  - SRAM column, 371
  - statistical distribution, 370
  - 2-D example, 361–363
- Scanning electron microscope (SEM), 24
- Selective boosting and write-assist circuitry
  - 14 nm FinFET SOI technology die, 340, 341
  - gate and source nodes, capacitive coupling, 338, 339
  - negative boost write assist, 339
  - programmable boost designs, 340
  - 6-transistor SRAM cell, 338
  - SRAM cell cross-section, 340
- Semisupervised learning, 2–3
- Shipped product quality loss (SPQL), 246
- Sigmoidal function, 474, 475
- Signal probabilities (SPs), 266–267
- Silicon characterization
  - chip-level, 220
  - cost, 202
  - issues, 202
  - Nyquist sampling theorem, 202–203
  - VP (*see* Virtual probe (VP))
  - wafer-level, 219–220
- SiLVR model, 12
  - complexity, 477
  - convergence, 477–479
  - hyperbolic tangent function, 474
  - implementation, 488–489
  - interpretations
    - input-referred correlation, 480–481
    - relative global sensitivity, 479–480
  - i*-th component, 475
  - i*-th ridge function, 476
  - linear projection, 475
  - master–slave flip-flop with scan chain, 490–492
  - nonlinear projection, 475
  - realistic circuit test, 489
  - representation, 473
  - sigmoidal function, 474, 475
  - sub-1 V CMOS bandgap voltage reference, 497–500
  - training algorithm, 476–477
    - Bayesian regularization, 486–487
    - cross-validation, 487–488
    - fast convergence, 482
    - good generalizability, 481
    - Levenberg–Marquardt algorithm, 483–486
    - robust convergence, 482
    - Spearman’s rank correlation, 482–483
- 2-LV, 488–489
- two-stage RC-compensated opamp
  - analog domain, 492, 493
  - gain, 495, 496
  - input offset, 495
  - observations, 494
  - performance metrics, 492
  - projection vector, 492, 493
  - quantitative comparisons, 496
  - quantitative measures, 495

- SiLVR model (*cont.*)
  - simulated and predicted values, 492, 494
  - $V_i$  standard deviation, 492
- Simulated annealing (SA) algorithm, 133
- Simulation traces, 375, 376
- Singular value decomposition (SVD), 47
- Smoothing method, 465–467
- Sparse relevance kernel machine (SRKM), 12
  - Bayesian learning model, 433–435
  - Bernoulli likelihood, 435
  - BG and IREF blocks, 443, 444
  - classification and regression problems, 425
  - expectation maximization (EM) algorithm, 436
  - process design kit, 443
  - relevance vector machine, 431–432
  - TSD block, 444
- Spatial correlation, 121, 238
- Spatial correlation function extraction, 239
- Spatiotemporal Gaussian process with progressive sampling (GP-ST-PS), 140, 142
- Spearman's rank correlation, 482–483
- Speed binning, 220–221
- Squared exponential (SE) kernel function, 171
- SqueezeNet, 657
- SRAFs, *see* Sub-resolution assist features
- SRAM bit-cell, 349, 350
- SSS, *see* Scaled-sigma sampling
- Statement coverage, assertion, 620
- State-of-the-art FinFET SRAM design
  - experimental setup, 341
  - importance sampling stage, 343, 344
  - selective boosting and write-assist circuitry, 338–341
  - uniform sampling stage, 342, 343
  - yield estimation and convergence analysis, 343–346
- Static timing analysis (STA), 285, 377
- Statistical blockade
  - algorithm, 309
  - blockade filtering, 308
  - circuit performance metric, 307
  - conditionals and disjoint tail regions, 313–316
  - efficient tail sampling method, 309, 310
  - extremely rare events and their statistics, 316–317
  - recursive formulation, 317–321
  - sampling and commercially available device models, 310–311
  - sampling method, 305
  - 6T SRAM Cell, 311–313
  - tail and body regions, 308
  - unbiasing classifier, 309
- Steiner tree wirelength (StWL), 107
- Step-by-step correlation, 284
- Stone–Weierstrass theorem, 461–462
- Stratified sampling, 327
- Sub-resolution assist features (SRAFs)
  - effectiveness of, 102–103
  - EPE, PV band and runtime, 104–105
  - logistic regression, 103–104
  - model-based generation, 102
  - rule-based generation, 102
  - supervised learning, 103
  - SVM, 104
- Subset simulation (SUS)
  - CLT, 357, 358
  - conditional PDFs, 351, 355, 356
  - confidence interval, 357, 361, 362
  - continuous PoI, 351
  - covariance, 360
  - efficacy, 360
  - failure rate of circuit, 352–353
  - “golden” failure rate, 360–361
  - intermediate failure events, 352
  - MCMC samples, 355–358
  - MM algorithm, 351, 355
  - normal distribution, 359
  - rare failure probability estimation, 351
  - rare failure rate, 357
  - SRAM column example, 360, 361
  - SSS approach, 351, 361
  - 2-D example, 353, 354
  - vector, 352
- Supervised learning, 2–3, 379
  - cost function, 36
    - feature-generation parameters, 45–46
    - feature vector, 44
    - functions, 44–45
    - linear regression parameters, 46–47
    - loss function, 43
    - measured and predicted CDs, 42–43
    - MSE<sub>train</sub>, 43
    - nonideal performance, 47
    - parameter vector, 43
    - regularization hyperparameters, 46
    - test set, 43
    - training set, 43
    - weight vector, 44
  - dataset, 36
    - input feature generation, 38–40
    - output representation, 38, 40–42
    - resist boundary contour, 37
  - definition, 36

- linear regression model form, 37
- optimization procedure, 36–37
- Support-vector machine (SVM), 98, 266, 423
  - predictor training using, 273–275
  - training and validation, 282
- Support vector regression (SVR), 71
- Surrogate metrics, 245
- SUS, *see* Subset simulation
- Synopsys Primitime, 285
- Synthesis tuning system (SynTunSys), 13
  - architecture, 541–543
  - archive, 542
  - automated decision loop, 541
  - components
    - cost function, 547
    - main tuning loop, 547
    - rules file, 542, 546
    - secondary tuning loop, 547
  - decision engine algorithms, 547–548
    - base decision algorithm, 549–550
    - comparisons, 556–559
    - 14 nm data mining, 555
    - IBM server chips, 554
    - learning algorithm, 551–553
    - z13 22 nm server processor., 554–555
  - DSE process, 539
  - footprint efficiency, 562–564
  - history-based enhancement
    - jump-start algorithm, 559–561
    - recommender system, 561–562
  - human design time vs. compute time, 542
  - parameters, 544
  - primitive library, 546
  - primitives, 542, 545
  - quality of results, 540
  - scheduler
    - priority ranking, 565–566
    - second pass SynTunSys runs, 566
    - STSS results, 566–567
  - synthesis parameters, 540
  - total negative slack, 541
  - VLSI chip design, 568

## T

### Tails

- and extreme values
  - Monte Carlo based methods, 300
  - order statistics, 301–302
  - quantiles, 302–304
- learning GPD parameters
  - maximum likelihood estimation, 305–306

- probability-weighted moment matching, 306–307
- statistical blockade sampling method, 305
- and maxima
  - conditional CDF, 296–297
  - distribution tail, 299–300
  - sample maximum, 297–299
  - skewed distribution, SRAM metric, 296
- Target metrics, 245
- Testbench, 375
- Test margin, 246
- Thermal shutdown (TSD) function, 443
- 3-layer perceptrons (3LP), 458
- Time series analysis, 235
- Time-to-digital converters (TDCs), 440
- Training dataset, 381
- Transverse electric (TE) wave, 51
- Transverse magnetic (TM) wave, 51
- Trapping/detrapping (TD) model, 267, 268
- True positive rate (TPR), 181
- Tuning policy, 404
- Two-stage Bayesian optimization (TSBO)
  - acquisition function, 514–515
  - algorithm progression times, 517
  - Euclidean distance, 513
  - fast exploration stage, 512
  - hierarchical partitioning scheme, 514
  - IMGPO and GP-UCB, 516
  - learning acquisition functions, 513
  - log distance to optima, 516
  - in MATLAB, 515, 516
  - pure exploration stage, 514
  - run time statistics, 518
- Type II maximum likelihood (ML-II), 133

## U

- Uncertainty quantification, 12–13
  - description, 526–527
  - sensitivity analysis, 533–534
  - storage model, 530–533
- Underfitting, 384
- Unsupervised learning technique, 2, 235, 379

## V

- Validation dataset, 381
- Valid spatial correlation function, 239
- Variance reduction method, 327
- VeSC-CoL vs. CART, 391–396
- Virtual metrology, 235

- Virtual probe (VP), 9
- applications
    - chip-level silicon characterization, 220
    - post-silicon tuning, 221
    - speed binning, 220–221
    - wafer-level silicon characterization, 219–220
  - implementation
    - advantages, 217–218
    - DCT coefficient preselection, 216–217
    - issues, 212–213
    - Latin hypercube sampling, 214–216
    - linear programming, 213–214
    - normalization, 213
  - MAP estimation, 203
    - accuracy of, 210–212
    - DCT coefficients, 206–207
    - Dirac delta function, 208
    - flush delay, 206
    - frequency domain, sparse pattern in, 204
    - likelihood function, 208, 209
    - $L_1$ -norm regularization, 209–210
    - MLE, 207
    - optimally fitted Laplace distribution, 207, 208
    - PDF, 207, 208
    - posterior distribution, 209
    - prior distribution, 206, 209
  - mathematical formulation, 204–206
  - numerical experiments
    - flush delay measurement data (*see* Flush delay measurement data)
    - leakage current measurement data, 226–228
    - ring oscillator period measurement data, 228–230
    - 2.8-GHz Linux server, 221
    - “smart” algorithm, 203, 208
    - test structures, 203, 204
    - wafer/chip, spatial variation of, 203
  - Virtual probe (VP) approach, 122
  - VLSI computer-aided design (VLSI CAD)
    - abstraction levels, 5–6
  - Voltage-controlled oscillator (VCO), 419–420, 440
  - Voltage regulation module (VRM), 579
- W**
- Wafer-level silicon characterization, 219–220
  - Weight sharing, 655–656
  - Within-die (WID) variation, 237
- X**
- Xception, 656
- Y**
- Yield enhancement, 8–9
- Z**
- Zach’s model, 41, 54–55
  - Zero-valued activations, 663
  - z13 22 nm server processor., 554–555